Klasifikasi Daya Tarik Konten Artikel Media Daring Dari Data Google Analytics Dengan C-FDT

Erlin Windia Ambarsari*)

Jurusan Informatika, Fakultas Teknik Dan Ilmu Komputer, Universitas Indraprasta PGRI, Jakarta Jl. Nangka No. 58 C (TB. Simatupang), Tanjung Barat, Jagakarsa, RT.5/RW.5, Jakarta Selatan, Jakarta, 12530, Indonesia email: erlin.windia@unindra.ac.id

Received: 30 Maret 2018; Revised: 8 Mei 2018; Accepted: 14 Mei 2018 Copyright ©2018 Politeknik Harapan Bersama Tegal. All rights reserved

Abstract - Information of article which had attractive contains as Trending Topics, although this is article hoax or not. The frequency of article's content which created by online media, it can be monitored by Google Analytics. One of the reasons for using Google Analytics is to understand the content of a site which leads to the change and behavior of behind the content. Google Analytics can be regarded as web analytics software with ease of installation. Classification of Google Analytics data with C-Fuzzy Decision Tree (C-FDT), aims to get the attraction of article content, which means having special attention from visitors and the article can be interesting or not, and observed whether C-FDT can recognize patterns from metric data Google Analytics. The purpose of this study is the results of FDT are expected to facilitate online media managers to analyze the content of articles and evaluate content groups tend to potentially gain traffic for getting promotional or marketing advertising as revenue from online media sites. The results obtained are C-FDT can recognize the pattern of Google Analytics metrics thus as facilitating the search of the article content into a simple form that is the reduction of attributes by grouping data with the same object and the data had Pruning. Online media managers can focus on certain attributes that have a big effect on Content Articles. However C-FDT is having trouble dealing with data sync due to system errors when retrieving data from Google Analytics. Therefore it is necessary to monitor data in time series.

Abstrak - Informasi dalam bentuk artikel yang mempunyai daya tarik tersendiri menjadi Trending Topics atau viral, walaupun artikel tersebut termasuk hoaks atau tidak. pengukuran frekuensi konten artikel yang dibuat oleh media daring dapat dipantau oleh Google Analytics. Salah satu alasan penggunaan Google Analytics adalah untuk memahami konten situs yang mengarah pada perubahan dan perilaku di balik konten. Google Analytics dapat dikatakan sebagai perangkat lunak analisis web dengan kemudahan instalasi. Klasifikasi data Google Analytics dengan C-Fuzzy Decision Tree (C-FDT) untuk mendapatkan daya tarik konten artikel, yang berarti mempunyai perhatian khusus dari pengunjung sehingga artikel tersebut dapat dikatakan menarik atau tidak, kemudian diamati apakah C-FDT dapat mengenali pola dari data metrik Google Analytics.

*) Corresponding author: Erlin Windia Ambarsari

Email: erlin.windia@unindra.ac.id

Tujuan dari penelitian ini adalah hasil dari FDT diharapkan dapat memudahkan pengelola media daring untuk menganalisis konten artikel.dan mengevaluasi kelompok konten yang kemungkinan berpotensi dapat mendulang trafik sebagai syarat untuk mendapatkan iklan promosi maupun pemasaran sebagai pendapatan dari situs media daring. Hasil yang didapatkan adalah C-FDT dapat mengenali pola dari metrik Google Analytics sehingga memudahkan pencarian konten artikel ke dalam bentuk yang sederhana yaitu adanya pengurangan atribut dengan cara mengkelompokkan data dengan objek yang sama dan adanya Pruning. Pengelola media daring dapat fokus dengan atribut tertentu saja yang berpengaruh besar dengan Daya Tarik Konten Artikel. Namun C-FDT mengalami kesulitan ketika menghadapi ketidaksinkron data karena kesalahan sistem saat penarikan data dari Google Analytics. Oleh sebab itu perlu dilakukan monitoring data secara time

Kata Kunci - Media Daring, Google Analytics, C-FDT, Konten Artikel

I. PENDAHULUAN

Perkembangan Informasi di Indonesia sangat pesat, terutama surat-surat kabar media elektronik berbasis daring yang memberikan informasi kepada warganet untuk mengetahui apa yang terjadi didalam negerinya sendiri. Informasi dalam bentuk artikel yang mempunyai daya tarik tersendiri menjadi *Trending Topic* atau viral, walaupun artikel tersebut termasuk hoaks atau tidak. Namun hal ini menjadi perhatian khusus bagi generasi muda sebagai lahan bisnis yang menjanjikan karena pesatnya perkembangan informasi masa kini dimana membuat media daring sudah sangat mudah.

Salah satu media daring yang berbasis M.I.C.E (*Meeting*, *Incentive*, *Convention*, *and Exhibition*) *Media Have Fun* menyajikan artikel bagi warganet dilihat dari sisi budaya, teknologi dan kegiatan anak-anak generasi muda dalam membuat acara dimana membuat konten artikel menarik dengan tujuan menaikan peringkat situs, sehingga dapat menarik pengunjung yang mau membaca konten artikel. Keuntungan didapatkan dapat berupa periklanan ataupun

undangan dari *Event Organizer* yang berujung pada bisnis pemasaran dan promosi.

Atas dasar keuntungan tersebut, pengukuran frekuensi konten artikel yang dibuat oleh media daring dapat dipantau oleh *Google Analytics*. Salah satu alasan penggunaan *Google Analytics* adalah untuk memahami konten situs yang mengarah pada perubahan dan perilaku di balik konten [1]. *Google Analytics* dapat dikatakan sebagai perangkat lunak analisis web dengan kemudahan instalasi [2].

Pertimbangan yang perlu diperhatikan, bahwa perubahan dan perilaku disebabkan daya tarik dari konten artikel yang berkualitas SEO (*Search Engine Optimization*) dengan tujuan dapat mendulang trafik sehingga mudah terindeks oleh mesin pencarian *Google*. Semakin konten artikel mudah diindeks, maka peringkat situs akan menaik, sehingga warganet akan sering mengunjungi situs tersebut.

Namun, permasalahan terjadi pada Google Analytics adalah sulitnya menganalisis konten artikel oleh pengelola media daring. Permasalahan lain yang terjadi, ketidaksinkron data metrik yang ditarik dari Google Analytics yang memungkinkan data tidak dapat dibaca. Karena itu diperlukan metode untuk menyederhanakan data dengan cara mengelompokan variabel yang memiliki kemiripan sehingga memudahkan pengelola media daring membaca indikator dari Google Analytics serta menangani data-data yang kemungkinan tidak dapat terbaca karena kendala sistem dari Google Analytics, sehingga metode yang digunakan sebagai alat bantu untuk penelitian ini adalah C-Fuzzy Decision Tree (C-FDT). C-FDT menjadi salah satu solusi untuk mengetahui pola data pengunjung berdasarkan setiap konten artikel, karena daya tarik konten bergantung kepada minat warganet membaca artikel.

Berdasarkan alasan diatas, penulis melakukan klasifikasi data *Google Analytics* dengan C-FDT untuk mendapatkan daya tarik konten artikel, yang berarti mempunyai perhatian khusus dari pengunjung sehingga artikel tersebut dapat dikatakan menarik atau tidak, kemudian diamati apakah C-FDT dapat mengenali pola dari data metrik *Google Analytics*. Tujuan dari penelitian ini adalah hasil dari FDT diharapkan dapat memudahkan pengelola media daring untuk menganalisis konten artikel.dan mengevaluasi kelompok konten yang kemungkinan berpotensi dapat mendulang trafik sebagai syarat untuk mendapatkan iklan promosi maupun pemasaran sebagai pendapatan dari situs media daring.

II. PENELITIAN YANG TERKAIT

Penelitian mengenai Google Analytics dari Omidvar M.A., Mirabi V.R, dan Shokry N membahas bagaimana warganet yang mengunjungi situs dapat menyebabkan dampak besar pada tampilan halaman. Google Analytics pada penelitian tersebut menggunakan autoregressive-moving-average (ARMA) models dengan mengambil 19.703 entri data selama 23 bulan untuk dianalisis kinerja tampilan halaman atau tampilan perhalaman yang dikunjungi [3]. Penelitian mengenai Google Analytics tidak berlanjut pada pengembangan sistem masa sekarang, sehingga penulis

mengangkat kembali masalah konten per halaman artikel dari banyaknya pengunjung tersebut untuk periode 2017-2018, dengan mencari tahu penyebab dampak besar pada tampilan halaman dilihat dari data tarik konten dengan ketidakpastian data dari metrik *Google Analytics*. Penggunaan *Fuzzy Decision Tree* (FDT) merupakan algoritma ditujukan untuk menangani ketidakpastian data. Algoritma FDT telah menunjukkan keunggulan algoritma pohon keputusan klasik [4]. Namun pada FDT yang berdasarkan teori himpunan *fuzzy* memiliki beberapa keterbatasan yang berasal dari bagian dalam dirinya [5].

Banyak cara mengkolaborasikan himpunan fuzzy dalam logika fuzzy dengan 2 atau 3 metode seperti : ANFIS[6], Fuzzy Ant[7], FMADM[8], Fuzzy C-Means[9]. Metode FDT dengan Fuzzy C-Means dapat dipertimbangkan untuk menangani ketidakpastian data Google Analytics, Fuzzy C-Means dalam FDT dikatakan sebagai C-Fuzzy Decicion Tree (C-FDT), dalam penelitian Gadomer L. dan Sosnowski Z.A mengatakan bahwa C-FDT yang diusulkan oleh Pedrycz W. dan Sosnowski Z.A [10] dimana gagasan jenis pohon ini didasari dengan asumsi memperlakukan data sebagai kumpulan butiran informasi yang hampir sama dengan kelompok Fuzzy. Pohon yang diusulkan dibentangkan diatas butiran-butiran tersebut. C-FDT mengasumsikan data pengelompokan dalam butiran yang dicirikan variabilitas rendah (yang berarti objek yang sama sampai ke cluster yang sama). Butiran ini adalah blok bangunan utama dari suatu pohon [11].

Pada dasarnya dalam membangun pohon FDT berbasis pada ID3 [12], [13], [14]. ID3 dilanjutkan oleh WEKA (Waikato Environment for Knowledge Analysis) dengan nama J48. J48 adalah implementasi Java dari algoritma C4.5 dengan melakukan perbaikan dari ID3, termasuk mengizinkan variabel kontinu menjadi fitur-fitur dengan memilih nilai ambang (threshold) dengan cara dipisahkan, menjadi lebih fleksibel melakukan membuat dan pemangkasan pohon (pruning tree) yang membuat klasifikasi menjadi efisien. Penelitian yang dilakukan Ahishakiye E., Taremwa D., Omulo E.O., dan Niyonzima I memilih J48 model klasifikasi dikarenakan kinerjanya mengadaptasi pada kumpulan data. Hasil eksperimen dari WEKA menunjukan bahwa J48 dapat memprediksi kategori data kriminal yang tidak diketahui dengan akurasi 94,25287% dimana sistem ini cukup baik dalam keandalannya untuk memprediksi kejahatan dan membutuhkan waktu sedikit untuk dijalankan dibandingkan dengan algoritma klasifikasi lain [15].

Berdasarkan pengumpulan literatur yang dijabarkan oleh penulis dalam penyelesaian masalah untuk *Google Analytics*, maka metode yang digunakan antara lain C-FDT dengan basis J48 (kolaborasi *Fuzzy C-Means* dan J48). Terdapat perbedaan dari [10] yaitu hasil *clustering* dari *Fuzzy C-Means* yang sudah di interferensi bagi objek yang sama dan data dengan objek yang berbeda, kemudian dimasukkan kedalam *Class* J48 sebagai *instance* termasuk didalamnya kategori data yang tidak diketahui atau data yang tidak dapat di baca

dari *Google Analytics* (pemodelannya akan dijelaskan pada sub bagian C-Fuzzy Decision Tree (C-FDT)).

III. METODE PENELITIAN

A. Fuzzy C-Means (FCM)

Metode *Clustering* (Pengelompokkan) *Fuzzy C-Means* yang dikembangkan oleh J.C Dunn tersebut bertujuan untuk mengenali pola dengan mengizinkan beberapa data yang dibagi ke dalam beberapa *cluster*, berikut langkah-langkah algoritma *Fuzzy* yang berdasarkan dengan konsep c-*partition* [7]:

1) $\{x_1,...,x_n\}$ merupakan himpunan data dari X, dimana setiap titik data dari xk (k=1,...,n) merupakan vektor dalam dimensi \Re^p , sehingga persamaan diasumsikan sebagai berikut [16]:

$$\begin{array}{l} M_{fcn} = \{U \in U_{cn} : u_{ik} \in [0,1] \sum_{i=1}^{c} u_{ik} = 1,0 < \\ \sum_{k=1}^{n} u_{ik} < n \} \end{array}$$

Dimana:

 U_{cn} adalah himpunan bilangan real dengan matrik c x n; c adalah integer dengan $2 \le c \le n$.

 U_{ik} adalah nilai anggota dengan x_k pada $\mathit{cluster}$ i (i=1,...,c).

2) Target dari FCM adalah menemukan pilihan *fuzzy* c-partition dan dibangun dalam bentuk minimalisasi fungsi objektif, seperti persamaan di bawah ini [16]:

$$J_m(U, V; X) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m ||x_k - v_i||^2$$
 (2)

Dimana:

 $V = (v_1, v_2,...,v_c)$ adalah matrik *cluster center* $v_i = \Re^p$. $\|.\|$ adalah *Euclidean norm*.

Bobot eksponen m dalam $[1, \infty)$ adalah konstan mempengaruhi nilai derajat keanggotaan.

- Kriteria minimalisasi J_m yang dikendalikan fuzzy, FCM didefinisikan sebagai algoritma minimalisasi bergilir yaitu:
 - a) Memilih nilai c, m, dan ε sebagai konstanta kecil positif.
 - b) Menghasilkan *fuzzy* c-*partition* secara acak U^0 dan mengatur iterasi t = 0.
- 4) Proses perulangan dua langkah tersebut berjalan dengan memberikan nilai derajat keanggotaan $u_{ik}^{(t)}$, *cluster center* $v_i^{(t)}$ (i =1,...,c) dihitung oleh [16]:

$$v_i^{(t)} = \frac{\sum_{k=1}^{n} (u_{ik}^{(t)})^m x_k}{\sum_{k=1}^{n} (u_{ik}^{(t)})^m}$$
(3)

5) Memberikan *cluster center* baru $v_i^{(t)}$ dengar memperbaharui nilai derajat keanggotaan $u_{ik}^{(t)}$ [16]:

$$v_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{\left\| x_k - v_i^{(t)} \right\|^2}{\left\| x_k - v_j^{(t)} \right\|^2} \right)^{\frac{2}{m-1}} \right]^{-1}$$
(4)

6) Proses akan berhenti ketika $|U^{(t+1)} - U^{(t)}| \le \varepsilon$ atau jumlah iterasi yang ditentukan sudah tercapai.

B. Algoritma J48

J48 merupakan kelanjutan dari algoritma ID3 dengan menambahkan fitur seperti menghitung nilai (*value*) yang hilang, pemangkasan *decision tree*, derivasi aturan (*rule*), rentang nilai atribut secara kontinu dan sebagainya. J48 juga merupakan implementasi dari algoritma C4.5 di Weka dimana klasifikasi dilakukan secara rekursif sampai setiap *leaf* tunggal menjadi jelas dengan menghasilkan aturan identitas data tertentu yang dikeluarkan. Tujuannya adalah mengeneralisasikan secara progresif dari *decision tree* sampai mendapatkan keseimbangan yang luwes dan akurat.

Langkah-langkah algoritma J48 adalah sebagai berikut:

- 1) Jika *Instance* termasuk *Class* yang sama dengan *Tree* yang mewakili *Leaf*, maka *Leaf* yang sudah diberi tanda dikembalikan ke dalam *Class* yang sama.
- 2) Informasi potensial dihitung untuk setiap atribut dan diberikan pengujian pada atribut, maka dihitung untuk mendapatkan hasil perolehan (*Gain*) informasi. Proses ini menggunakan *Entropy*, dimana formulanya sebagai berikut [17]:

$$Entropy(\vec{y}) = \sum_{j=1}^{n} \frac{|y_i|}{|\vec{y}|} log(\frac{|y_i|}{|\vec{y}|})$$
 (5)

$$Entropy(j|\vec{y}) = \frac{|y_j|}{|\vec{y}|} log\left(\frac{|y_j|}{|\vec{y}|}\right)$$
 (6)

Tujuannya untuk memaksimalkan *Gain* dengan membagi keseluruhan *Entropy* dengan memisahkan uraian \vec{y} dengan nilai j[17].

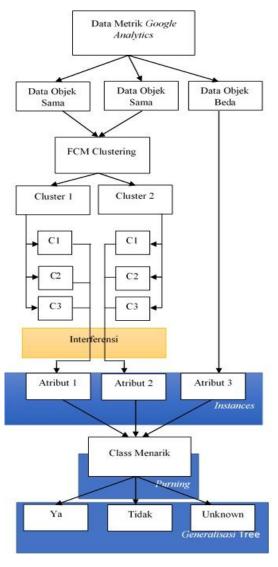
$$Gain(\vec{y},j) = Entropy(\vec{y} - Entropy(j|\vec{y}))$$
 (7)

- 3) Atribut terbaik didapatkan dari hasil seleksi dan dimasukkan ke dalam *Branch*.
- 4) Beberapa *Instance* dalam himpunan data (*data set*) kemungkinan tidak terdefinisi, karena itu diperlukan pemangkasan (*Pruning*). *Pruning* dilakukan untuk mengurangi kesalahan klasifikasi yang sedang diproduksi oleh spesialisasi dalam *Training Set. Pruning* dilakukan untuk generalisasi *Tree*.

C. C-Fuzzy Decision Tree (C-FDT)

C-FDT untuk mengklasifikasi data *Google Analytics* berdasarkan daya tarik konten artikel, dengan melakukan penarikan data metrik *Google Analytics*. Himpunan data metrik kemudian dikelompokkan dan dihitung dengan FCM untuk mendapatkan derajat keanggotaan tiap *cluster* bagi himpunan data dengan objek yang sama, sedangkan

himpunan data yang berbeda dipisahkan tersendiri. Kemudian tiap cluster dengan derajat keanggotaan tertinggi di interferensikan. Hasil interferensi tiap *cluster* dimasukkan sebagai atribut, termasuk himpunan data yang tadi dipisahkan. Penetapan *class* menarik atau tidak dilihat dari *instance* setiap data. Jika ada data yang tidak dapat dibaca dari metrik *Google Analytics*, maka diberi label *unknown* pada *class*. Lebih jelas mengenai tahap-tahap penelitian ini dalam pengolahan C-FDT dapat dilihat pada Gbr. 1.



Gbr 1. Tahapan Penelitian

IV. HASIL DAN PEMBAHASAN

Data yang diambil dari *Google Analytics* pada media daring *Media Have Fun* yang dilakukan selama 4 bulan dari bulan Januari 2017 sampai dengan Februari 2018 (Gbr. 2) dengan sebanyak 217 konten artikel, dimana rata-rata

warganet yang banyak mengunjungi situs tersebut adalah anak-anak muda berusia 19 tahun dari Indonesia (Gbr. 3).



Gbr. 2 Frekuensi Pengunjung Situs Bulan November 2017



Pengguna berdasarkan negara

Negara	Pengguna		
Total	6.438		
Indonesia	6.124		
United States	123		
Malaysia	35		
Japan	34		
Peru	32		

Gbr. 3 Pengunjung Berdasarkan Negara

TABEL I
DATA GOOGLE ANALYTICS MEDIA HAVE HUN

Halaman	Tayangan Halaman	Tayangan Halaman Unik	Waktu Rta. Di Halaman	Masuk	Rasio Pentalan	% Keluar	Nilai Halaman
/tahun-ini-starnifest-2017- hadir-kembali-di-surabaya	667	325	36.60	311	50.48%	43.78%	0.00
/japanese-film-festival- 2017-hadir-di-cgv-grand- indonesia	452	156	37.20	146	11.64%	27.65%	0.00
/japanese-film-festival- 2017-opening-ceremony	151	66	48.16	37	13.51%	27.15%	0.00
/escape-room-didalam- kampus-dengan-suasanya- menyeramkan	132	37	16.89	27	0.00%	12.88%	0.00

Data Google Analytics di tarik (ekspor) ke dalam Microsoft Excel, dimana data yang didapatkan adalah Nama Konten Artikel (Halaman), Tayangan Halaman, Waktu Rata-rata Halaman, Masuk, Rasio Pentalan, Keluar, dan Nilai Halaman, dilihat dari tabel 1. Data Google Analytics pada tabel 1 mempunyai objek yang sama, seperti Tayangan Halaman dan Tayangan Halaman Unik atau Rasio Pentalan dan %Keluar. Objek tersebut dibuat Clustering Fuzzy C-Means dibagi menjadi 2 atribut yang terdiri dari 2 sub-atribut. Data dari atribut Waktu Rata-rata Halaman dan Masuk dipisahkan karena bukan objek yang sama, sedangkan atribut Nilai Halaman di abaikan karena nilainya nol.

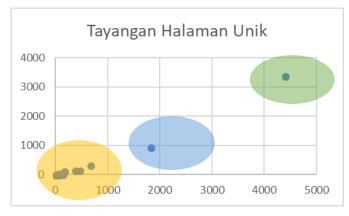
Nama atribut diasumsikan dengan nama atribut Tayangan yang terdiri dari sub-atribut Tayangan Halaman dan Tayangan Halaman Unik, atribut Pementalan terdiri dari sub-atribut Rasio Pentalan dan %Keluar. Kemudian masingmasing dari Kedua atribut dibuat *clustering* 2 dimensi dengan membagi menjadi 3 *cluster*, yaitu c₁, c₂, dan c₃. Nilai anggota tiap *cluster* dibuat secara acak yang kemudian dihitung fungsi obyektif dengan nilai bobot eksponen adalah 2. Nilai bobot ini dapat mempengaruhi nilai derajat keanggotaan.

Iterasi berhenti pada Iterasi ke 30 pada masing-masing atribut, dimana pada atribut Tayangan dengan nilai kesalahan 2.44E-09 dan atribut Pementalan dengan nilai kesalahan 3.54E-07 dengan nilai derajat keanggotaan pada tabel II.

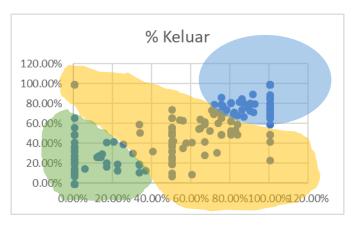
Kemudian nilai derajat keanggotaan terbesar diambil berdasarkan nilai *cluster* c₁, c₂, dan c₃, seperti pada tabel 2 bahwa data 1 sampai 19 semua berada di *cluster* c₃. Data-data *clustering* dapat dilihat pada grafik dibawah ini (Gbr 4 dan 5).

TABEL II Nilai Derajat Keanggotaan *Class* Tayangan

	LAI DERAJAT KEANG		
μ	C ₁	C 2	C 3
1	0.016106553	0.255197	0.728697
2	0.007180189	0.084858	0.907962
3	0.000662028	0.006002	0.993336
4	0.000425952	0.003779	0.995795
5	0.000383666	0.003386	0.996231
6	0.00028559	0.002496	0.997218
7	0.000115032	0.000981	0.998904
8	0.000160153	0.001377	0.998463
9	4.37181E-05	0.000368	0.999589
10	3.60524E-05	0.000303	0.999661
11	2.67426E-05	0.000224	0.99975
12	3.83983E-05	0.000322	0.999639
13	2.53258E-05	0.000212	0.999763
14	2.12384E-05	0.000177	0.999802
15	2.29703E-05	0.000192	0.999785
16	2.01724E-05	0.000168	0.999812
17	1.80418E-05	0.00015	0.999832
18	1.92855E-05	0.000161	0.99982
19	8.64481E-06	7.17E-05	0.99992



Gbr. 4 Clustering atribut Tayangan



Gbr. 5 Clustering atribut Pementalan

TABEL III INTERFERENSI DAN KLASIFIKASI ATRIBUT

	Waktu			
	Rata-rata			
Tayangan	Halaman	Masuk	Pementalan	Menarik
Rendah	36.60	311	Sedang	Tidak
Rendah	37.20	146	Rendah	Ya
Rendah	48.16	37	Rendah	Ya
Rendah	16.89	27	Rendah	Ya
Rendah	16.87	11	Rendah	Ya
Rendah	16.09	19	Rendah	Ya
Rendah	35.86	9	Rendah	Ya
Rendah	39.94	11	Sedang	Tidak
Rendah	19.98	9	Rendah	Ya

Hasil *Cluster* kemudian diinterferensikan menjadi Rendah, Sedang dan Tinggi yang diambil dari *cluster* c₁, c₂, dan c₃. Penentuan dapat dilihat dari Gbr. 4 dan 5; yaitu atribut Tayangan pada Gbr. 4 yang diberi warna hijau adalah Tinggi, warna biru adalah Sedang, dan warna kuning adalah Rendah. Atribut Pementalan diberi warna hijau adalah Rendah, warna kuning adalah Sedang, dan warna biru adalah Tinggi. Hasil yang sudah diinterferensikan dapat dilihat tabel III.

Masing-masing atribut, yaitu: atribut Tayangan mewakili konten yang berkualitas SEO, atribut Waktu Rata-rata Halaman mewakili lamanya warganet mengunjungi konten artikel, atribut Masuk mewakili banyaknya warganet yang masuk untuk melihat konten artikel, atribut Pementalan mewakili warganet yang keluar dari konten artikel, Atribut tersebut dijadikan satu dilihat *instance* dalam penentuan *Class* Menarik untuk mengetahui apakah konten artikel dapat dikatakan menarik atau tidak.

Namun kesulitan disini adalah ketidaksinkron data dari Google Analytics sehingga sulit untuk dianalisa dalam penentuan Class Menarik, misalkan terdapat nilai nol pada Waktu Rata-rata Halaman akan tetapi pengunjung yang masuk dikatakan banyak dan Pementalannya rendah, karena

itu penentuan dilakukan secara garis besar dimana dilihat Tayangan dari konten itu sendiri, yaitu semakin banyak yang masuk ke dalam konten dan Pementalan yang rendah dapat dikatakan artikel tersebut menarik. Bagi konten artikel yang tidak dapat didefinisikan menarik atau tidak, maka diberi label *unknown*.

Kemudian semua atribut dari Tabel 3 diolah ke dalam metode *Decision Tree* J48 menggunakan *software* WEKA dengan *Training Set* dimana didapatkan Klasifikasi *Instances* yang benar sekitar 93.5185 % sebanyak 202 data, sedangkan yang salah adalah 6.4815 % sebanyak 14 data pada Gbr. 6. Rincian data klasifikasi *Decision Tree* dapat dilihat pada Gbr. 7

Time taken to test model on training data: 0 seconds

=== Summary ===		
Correctly Classified Instances	202	93.5185 %
Incorrectly Classified Instances	14	6.4815 %
Kappa statistic	0.8512	
Mean absolute error	0.0787	
Root mean squared error	0.1983	
Relative absolute error	26.3843 %	
Root relative squared error	51.5391 %	
Total Number of Instances	216	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.967	0.076	0.967	0.967	0.967	0.891	0.955	0.965	Tidak
	0.983	0.057	0.864	0.983	0.919	0.891	0.970	0.865	Ya
	0.000	0.000	?	0.000	?	?	0.627	0.053	UnKnown
Weighted Avg.	0.935	0.068	?	0.935	?	?	0.947	0.905	

=== Confusion Matrix ===

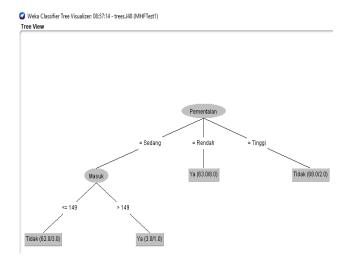
```
a b c <-- classified as

145 5 0 | a = Tidak

1 57 0 | b = Ya

4 4 0 | c = UnKnown
```

Gbr. 6 Training Set J48



Gbr. 7 Decision Tree

Pada gambar Gbr. 7 dapat dilihat bahwa ada 2 atribut dihilangkan, yaitu atribut Tayangan dan atribut Waktu Ratarata Halaman. Hilangnya kedua atribut disebabkan *Pruning* karena *Instance* dalam himpunan data dari atribut-atribut tersebut tidak terdefinisi, *Decision Tree* dari C-FDT dapat disimpulkan mengenai Daya Tarik Konten Artikel Media Daring adalah sebagai berikut:

- Jika Pementalan Tinggi maka konten artikel tidak menarik dengan kebenaran *instance* 88 data dan kesalahan *instance* 2 data.
- Jika Pementalan Rendah maka konten artikel menarik dengan kebenaran *instance* 63 data dan kesalahan *instance* 8 data.
- Jika Pementalan Sedang Dan Masuk <= 149 maka konten artikel tidak menarik dengan kebenaran instance
 62 data dan kesalahan instance 3 data.
- Jika Pementalan Sedang Dan Masuk > 149 maka konten artikel menarik dengan kebenaran *instance* 3 data dan kesalahan *instance* 1 data.

TABEL IV NAMA KONTEN YANG MENARIK

Nama Konten
/mengenal-kompetisi-ballroom-dance
/neo-comicfest-id-2017
/penampilan-perdana-samurai-x-otome-di-indonesia
/chibi-maruko-chan-meramaikan-jff-2017
/keseruan-beauty-workshop-japanese-kawaii-look-
bersama-canmake-x-reon-comics
/japanese-film-festival-2017-kini-hadir-kembali-di-empat- kota-di-indonesia
/lukisan-3d-kurumie-art-tutorialnya
/bermain-game-di-indonesia-comic-con-2017
/international-education-korean-contemporary-history- 2017
/jakarta-international-food-expo-jifex-creative-innovation-food-ingredients
/rudolf-seekor-kucing-hitam-kembali-hadir-dalam- japanese-film-festival-2017
/japanese-film-festival-2017-di-indonesia
/nitendo-switch-meranah-ke-penggemar-game-rpg- termasuk-skyrim
/lukisan-3d-kurumie-art-tutorialnya
/fakta-dari-cerita-greatest-showman

Berdasarkan pengolahan data yang telah dilakukan, C-FDT dapat mengenali pola dari metrik *Google Analytics* sehingga memudahkan pencarian konten artikel ke dalam bentuk yang sederhana yaitu adanya pengurangan atribut dengan cara mengkelompokkan data dengan objek yang sama dan adanya *Pruning*. Pengelola media daring dapat fokus dengan atribut tertentu saja yang berpengaruh besar dengan

Daya Tarik Konten Artikel. Namun C-FDT mengalami kesulitan ketika menghadapi ketidak-sinkron data karena kesalahan sistem saat penarikan data dari *Google Analytics*. Perlu dilakukan monitoring data secara *time series* sebelumnya seperti yang dilakukan oleh [3].

Berikut pada Tabel IV adalah artikel yang dinyatakan menarik berdasarkan hasil dari C-FDT. Pada tabel IV, dari hasil olahan C-FDT bahwa rata-rata warganet sering mengunjungi situs dengan konten artikel mengenai tema showbiz tentang perfilman yang ditayangkan serta korelasi dibalik *real story* pada film tersebut, tema game, tema kerajinan tangan, tema kuliner, serta acara jejepangan.

V. KESIMPULAN

Hasil penelitian yang dilakukan dapat disimpulkan bahwa data yang diolah berdasarkan *Training Set* pada J48 *Decision Tree* dimana Klasifikasi *Instances* yang benar sekitar 93.5185% sebanyak 202 data, sedangkan yang tidak benar adalah 6.4815% sebanyak 14 data. Konten Artikel Media Daring dapat dikatakan menarik adalah jika Pementalan Tinggi maka konten artikel tidak menarik, jika Pementalan Rendah maka konten artikel menarik, jika Pementalan Sedang Dan Masuk <= 149 maka konten artikel tidak menarik, jika Pementalan Sedang Dan Masuk > 149 maka konten artikel menarik.

C-FDT dapat mengenali pola dari metrik *Google Analytics* sehingga memudahkan pencarian konten artikel ke dalam bentuk yang sederhana yaitu adanya pengurangan atribut dengan cara mengkelompokkan data dengan objek yang sama dan adanya *Pruning*. C-FDT mengalami kesulitan ketika menghadapi ketidaksinkron data karena kesalahan sistem saat penarikan data dari *Google Analytics*. Berdasarkan data olahan C-FDT rata-rata warganet sering mengunjungi situs dengan konten artikel mengenai tema showbiz tentang perfilman tyang ditayangkan serta korelasi dibalik *real story* pada film tersebut, tema game, tema kerajinan tangan, tema kuliner, serta acara "jejepangan". Metode C-FDT perlu dikembangkan dengan penambahan metode time series khususnya untuk kasus pada *Google Analytics*.

UCAPAN TERIMA KASIH

Terimakasih kepada *Public Relationship Media Have Fun*, Gravita Alfiani untuk pengambilan data *Google Analytics*.

DAFTAR PUSTAKA

- [1] M. SOAT, "Google Analytics in 2016.," *Marketing News*, vol. 50, no. 2, pp. 8–9, 2016.
- [2] R. F. Uviña, "Bibliotecas y analítica web: una cuestión de privacidad," Información, cultura y sociedad, vol. 1, no. 33, pp. 105–112, Dec. 2015.
- [3] M. A. Omidvar, V. R. Mirabi, and N. Shokry, "ANALYZING THE IMPACT OF VISITORS ON PAGE VIEWS WITH GOOGLE ANALYTICS," *International Journal of Web & Semantic Technology (IJWesT)*, vol. 2, no. 1, pp. 14–32, 2011.
- [4] S. A. Ludwig, S. Picek, and D. Jakobovic, "Classification of Cancer

- Data: Analyzing Gene Expression Data Using a Fuzzy Decision Tree Algorithm," in *Operations Research Applications in Health Care Management*, Springer, Cham, 2018, pp. 327–347.
- [5] L. Van Tuong Lan, N. Mau Han, and N. Cong Hao, "AN ALGORITHM TO BUILDING A FUZZY DECISION TREE FOR DATA CLASSIFICATION PROBLEM BASED ON THE FUZZINESS INTERVALS MATCHING," Annales Univ. Sci. Budapest., Sect. Comp., vol. 45, pp. 319–335, 2016.
- [6] I. Afriliana, E. Budihartono, and A. Maulana, "Pengukuran Kinerja Dosen Menggunakan Metode Adaptive Neuro Fuzzy Inference System (ANFIS)," JURNAL INFORMATIKA: Jurnal Pengembangan IT, vol. 2, no. 2, pp. 109–112, Jul. 2017.
- [7] E. W. Ambarsari and S. Khotijah, "Clustering Algoritma Fuzzy Ant Untuk Optimalisasi Penentuan Rute Kemacetan Tanah Abang," Computatio: Journal of Computer Science and Information Systems, vol. 1, no. 2, pp. 187–194, Nov. 2017.
- [8] A. Hadiana, "Design of Kansei Engineering Support System Using Fuzzy Multiple Attribute Decision Making," *International Journal of Basic and Applied Science*, vol. 5, no. 1, pp. 24–30, 2016.
- [9] A. Stetco, X. J. Zeng, and J. Keane, "Fuzzy C-means++: Fuzzy C-means with effective seeding initialization," Expert Systems with Applications, vol. 42, no. 21, pp. 7541–7548, 2015.
- [10] W. Pedrycz and Z. A. Sosnowski, "C-fuzzy decision trees," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 35, no. 4, pp. 498–511, 2005.

- [11] Ł. Gadomer and Z. A. Sosnowski, "Knowledge aggregation in decision-making process with C-fuzzy random forest using OWA operators," *Soft Computing*, pp. 1–15, Jan. 2018.
- [12] J. Yun, J. Won Seo, and T. Yoon, "FUZZY DECISION TREE," International Journal of Fuzzy Logic Systems (IJFLS), vol. 4, no. 3, 2014.
- [13] S. J. Narayanan, I. Paramasivam, and R. B. Bhatt, "On the estimation of optimal number of clusters for the induction of fuzzy decision trees," *International Journal of Data Science*, vol. 2, no. 3, p. 221, 2017.
- [14] S. M. S. Askari and M. A. Hussain, "Credit card fraud detection using fuzzy ID3," in 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 446–452.
- [15] E. Ahishakiye, D. Taremwa, E. O. Omulo, and I. Niyonzima, "Crime Prediction Using Decision Tree (J48) Classification Algorithm," *International Journal of Computer and Information Technology*, vol. 06, no. 03, pp. 2279–764, 2017.
- [16] D. Parwatiningtyas, E. W. Ambarsari, and S. Mariko, "The calculation of the highest leak level of water pipe lines region at PDAM Tirta Kahuripan using fuzzy C-means and ArcGIS method analysis," AIP Conference Proceedings, vol. 030196, no. 1, p. 030196, 2017.
- [17] G. Kaur and A. Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes," *International Journal of Computer Applications*, vol. 98, no. 22, pp. 13–17, Jul. 2014.